Application of teh Kalman Filter to Interest Rate Modelling

Ibhagui, Oyakhilome

10 July 2010

Application of the Kalman Filter to Interest Rate Modelling

Oyakhilome Wallace Ibhagui (wallace@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by: Emeritus Prof. Ronnie Becker
University of Cape Town, South Africa

20 May 2010

Submitted in partial fulfillment of a postgraduate diploma at AIMS

Abstract

We give a mild introduction to the Kalman filter and the generalized Vasicek models of the term structure of interest rates with special attention to the application of the Kalman filter equations to one-and two-factor models. After thoroughly reviewing the essential tools that constitute the Kalman filter and the generalized Vasicek models of the term structure of interest rates, we derive the yield on a zero coupon bond with infinite maturity and the Kalman filter equations of the state space formulation of the generalized Vasicek models. By performing simulations, we illustrate how the Kalman filter works and the major weakness of the Vasicek model.

Wetin we do

Na small small we take introduce Kalman filter and generalized Vasicek models of the term structure of interest rates, where we look well well how we fit take apply Kalman filter equations to one-and two-factor models. After we don sabi everything wey dem dey take do Kalman filter and generalized Vasicek models, we come find the extra money wey person fit get ontop of the money wey another person borrow from am, when the time wey the person wey borrow the money go pay back go tay well well, and we come find the Kalman filter equations of the generalized Vasicek models after we don first write the generalized Vasicek models for one kind way wey dem dey call state space form. We come use computer take show how Kalman filter dey take sabi work and wetin make Vasicek model no too good. All dis thing wey I don talk na about money money matter and how we fit take make extra awoof money from the money wey somebody borrow from us. Oyaks Investment.

Declaration

I, the undersigned, hereby declare that the work contained in this essay is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Oyakhilome Wallace Ibhagui, 20 May 2010

Contents

1. Introduction

We start with a discussion of some important concepts in interest rate modelling.

1.1 Terms Associated With Interest Rates

Debt Instruments A debt Instrument is a promissory note that evidences a debtor/creditor relationship. In such a relationship, one party borrows funds from another party. The borrowing party promises to repay the borrowed funds together with interest on each unit of the borrowed funds. The borrowing party is the debtor and the lending party is the creditor. The promissory note is satisfied when the borrowing party's obligations of repaying the creditor with interest on the borrowed funds have all been met. A debt instrument may be marketable or non marketable. Marketable debt instruments are considered as securities. When the debt instrument takes the form of a security, that is when it is marketable, the borrower is called the issuer. The issuer sells its securities to the lender who is called an investor or holder. A bond is an example of a marketable debt instrument.

Bonds A bond is a financial security in which the holder is promised by the issuer a sequence of guaranteed future payments without any risk. It is risk free because the payments will be delivered with some degree of certainty. Nevertheless, risk is completely ineluctable since the market prices of bonds fluctuate unpredictably. The issuer of a bond sells it to the investor in order to raise money to finance other capital investments such as expanding into new markets; the issuer is required to pay the investor a fixed sum annually or semi-annually until maturity and then a fixed sum to repay the principal. Some bonds provide for periodic payments that include both interest and principal. Others only require periodic payments of interest. Still others require no payments of either interest or principal until the bond matures. When a bond requires that each payment includes some principal as well as interest in such a way that the principal is gradually repaid over the life of the bond, the bond is said to be amortizing. When a bond provides for the full repayment of principal in a lump sum at maturity, with interim payments limited to only the fixed rate of interest, the bond is said to be non-amortizing. Coupon bonds are non-amortizing.

Coupon Bonds Most debt instruments call for a fixed rate of interest that is paid periodically: semi-annually or annually, for example. The fixed rate of interest, which is always stated on an annual basis, is called the coupon rate and the payment itself is called the coupon. It is the interest rate that a bond issuer will pay to a bondholder. The coupon or coupon rate of a bond is the amount of interest paid per year expressed as a percentage of the bond's par value, which is the amount that the issuer is willing to pay the holder at maturity. This amount is unrelated to the market value of the bond. Coupon bonds are therefore a type of bond issue that offers a fixed rate of interest that is paid on a more frequent basis, with the par value of the bond paid in full at the time that the bond reaches maturity. One obvious advantage of a coupon bond is that it gives a regular source of revenue to its holder during a given calendar year. This leads us to a prominent class of coupon bonds in the financial market known as the zero coupon bonds. These bonds will be our main focus in this essay.

Zero Coupon Bonds When no payments of any kind are required until a bond matures, the bond is called a zero coupon bond or, more simply, a zero. Zero coupon bonds do not make regular interest payments like other bonds do. One receives all the interest payments in a lump sum at maturity. There is another advantageous feature of zero coupon bonds that should be noted. When the price of a bond is equal to its par value, the bond is said to be priced at par. When the price is greater than the par value, it is said to be priced at a premium. The price of a zero coupon bond is less than its par value. Hence we say a zero coupon bond is priced at a discount. All zero coupon bonds are always priced at a discount and so they are known as pure discount bonds. We can therefore use the words zero coupon bonds and pure discount bonds interchangeably. There is a great advantage attached to all zero coupon bonds: They have an inherent pure profit. The holders of zero coupon bonds are sure of getting more than what they invested, in addition to the fixed rate of interest, at maturity.

Term Structure of Interest Rates Interest rate is the financial term often used to describe the rate by which money is taken forward in time. The price one pays for the use of a unit of another's money is called an interest rate. This is exactly a bond's yield, which is the overall rate of interest that the issuer of a bond pays to the holder at maturity for the use of each unit of the holder's investment. The length of time until a bond matures is called its term to maturity. The general definition of yield is the return an investor will receive by holding a bond to maturity. For example, when the issuer of a bond receives payment from the holder, he is under an obligation to pay a price or interest on each unit of the holder's money at maturity. This payment is known as the yield. The concept of yield is one of the most important concepts in fixed-income security(bond) analysis. It is important to understand that different bonds have different yields. The relationship between yield and maturity is called the term structure of interest rates. When graphed, the term structure is called a yield curve. Yield curves give a measure of the market's expectations of future interest rates given the current market conditions.

1.2 Historical Overview

Since the last three decades, there has been a number of models of the term structure of interest rates. Recent evaluation of the term structure of various interest rate models has concentrated on the dynamic implication of the models using time-series or cross-sectional approach. A major set back of both approaches is that they do not use the full information obtained over time from the yield curve and across maturities in the estimation procedure [BN99]. More recent research has involved time-series and cross-sectional data using Kalman filtering methods. The application of Kalman filtering methods in the estimation of term structure models using cross-sectional and time-series data has been investigated by Pennachi (1991), Lund (1994), (1997), Chen and Scott (1995), Duan and Simonato (1995), Geyer and Pichler (1996), Ball and Torous (1996), and Jegadeesh and Pennachi (1996) [BN99]. Developed in 1960 by Rudolph Kalman, the Kalman filter is a set of mathematical equations that provide an efficient recursive computational technique for optimally estimating the state of an unobservable process in a way that minimizes the MSE. The filter is very powerful in several aspects: it supports estimations of past, present and even future states and it does all these even when the precise nature of the modelled system is unknown [WB06]. The term structure models, whose Kalman filter equations we shall consider, are the generalized Vasicek models of the term structure of interest rates. This application of the Kalman filter equations is based on the recent work of Babbs and Nowman [BN98] where the term structure models are written in a state space form to allow for measurement errors and the Kalman filter equations are used to provide the estimates of the unobserved state variables.

The objectives of this essay are: to provide a simple introduction to the theory of Kalman filter technique, to present the basic theory of the generalized Vasicek models of the term structure of interest rates and to apply Kalman filter technique to the generalized Vasicek models of the term structure of interest rates. As the time required for this essay is inevitably limited, the necessary implementation of this technique on actual bond markets will not be considered, but will certainly be accomplished in the near future. The structure of the rest of the essay is as follows: Chapter 2 discusses the general state space formulation, presents the full Kalman filter equations and simulations to illustrate how the equations work. Chapter 3 explains the generalized Vasicek models of the term structure of interest rates, presents the general price of a zero coupon bond, offers the derivation of the yield on a zero coupon bond with infinite maturity and illustrates the simulation of the Vasicek model to uncover one of the model's major deficiencies. Chapter 4 applies the results obtained in Chapter 2 to write down the complete Kalman filter equations associated with the models studied in chapter 3. The last chapter offers some conclusions and areas of future research.

2. State Space Models and the Kalman Filter

The state space form is an indispensable tool that makes it possible to successfully handle an array of models. A model is said to be in a state space form if it is completely specified by two basic equations. These two equations are known as the measurement and transition equations. Once a model has been written in a state space form, the Kalman filter may be applied. The state space formulation is described in the first section of this chapter, while the second section develops the Kalman filter. Some examples are then given in the last section to illustrate how the Kalman filter works.

2.1 State Space Formulation

The general state space form applies to a multivariate time series, $\mathbf{y}_t$, containing N elements, where $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})$. The time series model can be represented via a state space model comprising two equations: measurement and transition equations. It should be noted that we are considering the case where the state space model is linear; that is, the observations in the measurement equation are a linear function of the state vector and, in the transition equation, the state vector is itself a linear function of the state vector in the previous time period and the matrix R_t does not depend on the lagged state vector. An example of a non-linear state space model is when the measurement and transition equations are given by

$$\begin{aligned} y_t &= Z_t\alpha_t^3 + d_t + \epsilon_t, \quad t = 1, \dots, T \\ \alpha_t &= \sqrt{T_t\alpha_{t-1} + c_t + R_t\eta_t}, \quad t = 1, \dots, T \end{aligned}$$

We will not consider this type of model because of its non-linearity.

2.1.1 Measurement Equation

The measurement equation specifies the observable variables and gives the relationship between the observable variables, $\mathbf{y}_t$, and the state vector, α_t, according to the equation

$$\mathbf{y}_t = Z_t\alpha_t + d_t + \epsilon_t, \quad t = 1, \dots, T, \tag{2.1}$$

where $\mathbf{y}_t$ is the observation process containing N elements, α_t is an $m \times 1$ vector, Z_t is an $N \times m$ matrix, d_t is an $N \times 1$ vector and ϵ_t is an $N \times 1$ vector known as the measurement noise. The measurement noise is the error associated with the measurement equation. The state vector contains the unobserved state variables, which are the variables to be estimated. To estimate the unobserved state variables, we must use the information provided by the observable variables in the measurement equation.

2.1.2 Transition Equation

In general, the elements of the state vector are not observable. They are governed by a first-order Markov process. The governing relationship is given by

$$\alpha_t = T_t\alpha_{t-1} + c_t + R_t\eta_t, \quad t = 1, \dots, T, \tag{2.2}$$

where T_t is an $m \times m$ matrix, c_t is an $m \times 1$ vector, R_t is an $m \times g$ matrix and η_t is a $g \times 1$ vector called the process noise. This is known as the transition equation.

The state vector denotes the unobservable vector that can be thought of as a vector that contains the necessary information to predict the observations. The inclusion of the matrix R_t in front of the process noise is somewhat arbitrary. The matrices Z_t, d_t and H_t in the measurement equation and the matrices T_t, c_t, R_t and Q_t in the transition equation are known as the system matrices. They contain the unknown parameters of the model.

When a model is written in a state space form, the choice of α_t is determined by construction. It is in general not unique. The aim of the state space formulation is to set up α_t in such a way that it contains all the relevant information on the system at time t. This is achieved when α_t has as small a number of elements as possible. A state space form in which the length of the state vector is minimized is called a minimal realization. A minimal realization is a basic criterion for a good state space representation. Again, however, this does not mean that there is necessarily a unique representation for any particular problem.

2.1.3 Assumptions on the State Space Model

1. The measurement errors are i) uncorrelated ii) additive and iii) normally distributed with mean zero and covariance matrix H_t, that is

   $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = H_t$.

2. The process disturbances are i) uncorrelated ii) additive and iii) normally distributed with mean zero and covariance matrix Q_t, that is

   $E(\eta_t) = 0$ and $Var(\eta_t) = Q_t$.

3. The measurement errors and the process disturbances are uncorrelated with each other in all time periods and uncorrelated with the initial state, that is

   $E(\epsilon_t \eta_s') = 0$ for all s, t=1, . . . ,T

   and

   $E(\epsilon_t \alpha_0') = 0, \quad E(\eta_t \alpha_0') = 0$ for all $t = 1$, . . . ,T.

4. The initial state vector, α_0, has a mean of a_0 and covariance matrix P_0, that is

   $E(\alpha_0) = \mathbf{a}_0$ and $Var(\alpha_0) = \mathbf{P}_0$.

2.1.4 Some Examples of State Space Formulation

In this section, we illustrate how to write a model in a state space form with some useful examples.

1. Consider the moving average MA(1) model

   $$y_t = p_t + \theta p_{t-1}.$$

   For this model, the transition and measurement equations are obtained by defining the state vector $\alpha_t = \begin{bmatrix} p_t \\ p_{t-1} \end{bmatrix}$ and putting it in the form (2.1) and (2.2). This gives

$\alpha_t = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} p_t$

and

$y_t = \begin{bmatrix} 0 & \theta \end{bmatrix} \begin{bmatrix} p_t \\ p_{t-1} \end{bmatrix} + p_t.$

Alternatively, the above MA(1) model can be put in a state space form by defining the state vector $\alpha_t = \begin{bmatrix} y_t \\ \theta p_t \end{bmatrix}$ and writing the transition and measurement equations as

$\alpha_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ \theta \end{bmatrix} p_t$

and

$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t, \quad t = 1, \ldots, T$

respectively. A feature of this representation is that there is no measurement equation noise. Moreover, the representation confirms the fact that the state space representation is in general not unique.

2. Consider the second-order autoregressive AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \quad t = 1, \ldots, T.$$

Two state space representations for this model are possible. First, we write the state vector as $\alpha_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}$ and then use (2.1) and (2.2) to obtain

$$\begin{aligned} y_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t, \quad t = 1, \ldots, T \\ \alpha_t &= \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t. \end{aligned}$$

Also, we could write the state vector as $\alpha_t = \begin{bmatrix} y_t \\ \phi_2 y_{t-1} \end{bmatrix}$ and use (2.2) and (2.1) to obtain

$$\begin{aligned} y_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \alpha_t, \quad t = 1, \ldots, T \\ \alpha_t &= \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t. \end{aligned}$$

The two examples show that the state space representation is not in general unique. As seen in both examples, the system matrices contain the unknown parameters of the models. For instance, in the AR(2) model, the unknown parameters are ϕ_1 and ϕ_2 and they are contained in at least one of the system matrices. One of the major statistical tasks will be to estimate these unknown parameters. This estimation is one of the goals that the Kalman filter accomplishes.

2.2 The Kalman Filter

When a model has been properly put in a state space form, a good technique for optimally estimating the unobservable state variables and the parameters of the model can now be applied. This prominent technique is known as the Kalman filter. The Kalman filter is essentially a set of mathematical equations that implement a recursive procedure for computing the optimal estimator of a state vector at time t, based on the information available at time t. It gives the optimal estimator of a state vector in the sense that it minimizes the covariance matrix of the estimation error, provided that the disturbances and initial state vector are normally distributed. The optimal estimator of a state vector given information available at time t is the conditional expectation of the state vector given information available at that same time. When the normality assumption is dropped, there is no longer any guarantee that the Kalman filter will give the conditional expectation of the state vector. However, it still gives the optimal estimator of the state vector in that it minimizes the covariance matrix of the estimation error.

2.2.1 General Form of the Kalman Filter Equations

Consider the state space model of (2.2) and (2.1). Let $\mathbf{a}_{t-1}$ denote the optimal estimator of α_{t-1} based on the observations up to and including $\mathbf{y}_{t-1}$ and let $\mathbf{P}_{t-1}$ denote the $m \times m$ covariance matrix of the estimation error, i.e.

$$\mathbf{P}_{t-1} = E_{t-1}\left[(\alpha_{t-1} - \mathbf{a}_{t-1})(\alpha_{t-1} - \mathbf{a}_{t-1})'\right]. \tag{2.3}$$

Let $\mathbf{a}_{t|t-1}$ denote the optimal estimator of α_t based on the observations up to and including $\mathbf{y}_{t-1}$ and let $\mathbf{P}_{t|t-1}$ denote the associated covariance matrix of the estimation error. According to the Kalman filter theory, the optimal estimator of α_t given information available at the immediate previous time is the conditional expectation of α_t based on the information available at the said time. So, $\mathbf{a}_{t|t-1} = E_{t-1}(\alpha_t)$.

Consequently, given $\mathbf{a}_{t-1}$ and $\mathbf{P}_{t-1}$, the optimal estimator of α_t is given by

$$\mathbf{a}_{t|t-1} = T_t\mathbf{a}_{t-1} + c_t, \tag{2.4}$$

while the covariance matrix of the estimation error is

$$\mathbf{P}_{t|t-1} = T_t\mathbf{P}_{t-1}T_t' + R_tQ_tR_t', \quad t = 1, \ldots, T. \tag{2.5}$$

These two equations are known as the prediction equations. They help in optimally estimating the unobserved state variables contained in the state vector at time t given information available at time $t-1$.

Once the new observation, $\mathbf{y}_t$, becomes available, the optimal estimator of α_t based on the information available at the immediate previous time can be updated to give the optimal estimator of α_t based on the information available at the current time t. That is, $\mathbf{a}_{t|t-1}$ can be updated to give $\mathbf{a}_{t|t} = \mathbf{a}_t$ and $\mathbf{P}_{t|t-1}$ can be updated to give $\mathbf{P}_{t|t} = \mathbf{P}_t$. This is called updating. The updating equations are

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}Z_t'F_t^{-1}(\mathbf{y}_t - Z_t\mathbf{a}_{t|t-1} - d_t) \tag{2.6}$$

and

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}Z_t'F_t^{-1}Z_t\mathbf{P}_{t|t-1}, \tag{2.7}$$

where

$$F_t = Z_t \mathbf{P}_{t|t-1} Z_t' + H_t, \quad t = 1, \ldots, T. \tag{2.8}$$

We assume that F_t^{-1} exists.

We can also get the optimal estimator at the next time step based on all the available information up to the current time step by using the updating equations. Doing so, we get

$$\begin{aligned}
\mathbf{a}_{t+1|t} &= T_{t+1}\mathbf{a}_t + c_{t+1} \\
&= T_{t+1}\left(\mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}Z_t'F_t^{-1}(\mathbf{y}_t - Z_t\mathbf{a}_{t|t-1} - d_t)\right) + c_{t+1} \\
&= T_{t+1}\mathbf{a}_{t|t-1} + T_{t+1}\mathbf{P}_{t|t-1}Z_t'F_t^{-1}(\mathbf{y}_t - Z_t\mathbf{a}_{t|t-1} - d_t) + c_{t+1} \\
&= T_{t+1}\mathbf{a}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - Z_t\mathbf{a}_{t|t-1} - d_t) + c_{t+1}.
\end{aligned}$$

So,

$$\mathbf{a}_{t+1|t} = (T_{t+1} - \mathbf{K}_t Z_t)\mathbf{a}_{t|t-1} + \mathbf{K}_t\mathbf{y}_t + (c_{t+1} - \mathbf{K}_t d_t), \tag{2.9}$$

where $\mathbf{K}_t$ is the Kalman gain matrix given by

$$\mathbf{K}_t = T_{t+1}\mathbf{P}_{t|t-1}Z_t'F_t^{-1}, \quad t = 1, \ldots, T. \tag{2.10}$$

Similarly, the recursion for the covariance matrix of the estimation error is obtained from (2.5) and is given by

$$\mathbf{P}_{t+1|t} = T_{t+1}(\mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}Z_t'F_t^{-1}Z_t\mathbf{P}_{t|t-1})T_{t+1}' + R_{t+1}Q_{t+1}R_{t+1}', \quad t = 1, \ldots, T, \tag{2.11}$$

which is known as a Ricatti equation.

We may specify the starting values for the Kalman filter in terms of $\mathbf{a}_0$ and $\mathbf{P}_0$ or $\mathbf{a}_{1|0}$ and $\mathbf{P}_{1|0}$. With these initial conditions, the Kalman filter gives the optimal estimator of the state vector as each new observation becomes available. When all observations have been processed, the filter delivers the optimal estimator of the current or next state vector based on the full information set. It is this estimator that contains all the information required to make the best predictions of future values of the state vector as well as the observations.

2.2.2 Properties of the Filter

After thoroughly studying the Kalman filter and its equations, it is important to understand the main characteristics and properties that make the filter an appropriate prediction tool. The Kalman filter is optimal with respect to virtually any criterion that makes sense. One aspect of this optimality, according to Maybeck, P. S. (1979), is that the Kalman filter incorporates all information that can be provided to it so that it combines all available measurement data, plus prior knowledge about the system and measurement devices, to produce an estimate of the desired variable in such a way that the error is minimized statistically. The mean, mode, median and any reasonable choice for an optimal estimate all coincide when the system can be described through a linear state space model and the system and measurement noises are white and Gaussian (whiteness implies that the noise value is not correlated in time, so that if you know what the value of the noise is now, this knowledge does you no good in predicting what its value will be at any other time [Gar06].

2.3 Derivation of the Kalman Filter

In this section, we will derive two of the five Kalman filter equations. Recall that the five Kalman filter equations are

1. State Prediction

$$\mathbf{a}_{t|t-1} = T_t\mathbf{a}_{t-1} + c_t$$

2. Covariance Prediction

$$\mathbf{P}_{t|t-1} = T_t\mathbf{P}_{t-1}T_t' + R_tQ_tR_t'$$

3. Kalman Gain

$$\mathbf{K}_t = T_{t+1}\mathbf{P}_{t|t-1}Z_t'F_t^{-1}$$

4. State Update

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}Z_t'F_t^{-1}(\mathbf{y}_t - Z_t\mathbf{a}_{t|t-1} - d_t)$$

5. Covariance Update

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}Z_t'F_t^{-1}Z_t\mathbf{P}_{t|t-1}$$

Now, taking expectation of both sides of (2.1) conditioned on the information available at time $t-1$ and using the state space assumption, we get

$$\begin{aligned}
E_{t-1}[\alpha_t] &= E_{t-1}[T_t\alpha_{t-1} + c_t + R_t\eta_t] \\
E_{t-1}[\alpha_t] &= T_tE_{t-1}[\alpha_{t-1}] + c_t \\
\mathbf{a}_{t|t-1} &= T_t\mathbf{a}_{t-1} + c_t,
\end{aligned}$$

which is the state prediction equation as written in the first equation.

For the second equation, we use the fact that

$$\mathbf{P}_{t|t-1} = E_{t-1}\left[(\alpha_t - \mathbf{a}_{t|t-1})(\alpha_t - \mathbf{a}_{t|t-1})'\right],$$

which, together with (2.2) and (2.4), gives

$$\begin{aligned}
(\alpha_t - \mathbf{a}_{t|t-1})(\alpha_t - \mathbf{a}_{t|t-1})' &= (T_t(\alpha_{t-1} - \mathbf{a}_{t-1}) + R_t\eta_t)\left((\alpha_{t-1} - \mathbf{a}_{t-1})'T_t' + \eta_t'R_t'\right) \\
E_{t-1}\left[(\alpha_t - \mathbf{a}_{t|t-1})(\alpha_t - \mathbf{a}_{t|t-1})'\right] &= T_tE_{t-1}\left[(\alpha_{t-1} - \mathbf{a}_{t-1})(\alpha_{t-1} - \mathbf{a}_{t-1})'\right]T_t' + R_tE_{t-1}\left[\eta_t\eta_t'\right]R_t'.
\end{aligned}$$

Consequently, from (2.5) and the state space assumptions, we get

$$\mathbf{P}_{t|t-1} = T_t\mathbf{P}_{t-1}T_t' + R_tQ_tR_t',$$

where $E_{t-1}\left[\eta_t\eta_t'\right] = Q_t$ is the covariance matrix of η_t.

The derivation of the last three equations reposes on the results given by the first two equations whose proofs we have given. Interested readers should see [Har89], [WB06] and [Gar06] for the complete proofs of the remaining Kalman filter equations.

2.4 Smoothing

Predicting and updating constitute the Kalman filter. Predicting involves finding an optimal estimator of a state vector at time t given information available up to time $t-1$, while updating involves finding an optimal estimator of a state vector at time t based on the information available up to that time. In the previous section, we saw that the Kalman filter helps in predicting and estimating. One other very important thing that it facilitates is smoothing. Although smoothing is not necessarily a filtering process, its recursion depends on the results that are available through the Kalman filter equations. This section gives a light introduction to the theory of smoothing.

The aim of smoothing is to take account of the information available after time t. The mean of the distribution of α_t, conditional on all sample available after time t, may now be written as $E(\alpha_t|I_T)$ and is known as a smoothed estimate. The corresponding estimator is called a smoother. Since the smoother is based on more information than the filtered estimator, it will have a covariance matrix of estimation error which, in general, is much less than that of the filtered estimator.

In a linear state space model, the smoothed estimator of α_t given information available after t, i.e. at T, is, as before, the conditional expectation of α_t given information available at time T. Let the smoothed estimator be denoted by $\mathbf{a}_{t|T}$, then

$$\mathbf{a}_{t|T} = E(\alpha_t|I_T) = E_T(\alpha_t).$$

As with the filtered estimator, $\mathbf{a}_{t|T}$ is the optimal estimator of α_t based on the available information set after time t. When the normality assumption is dropped, the smoothed estimator gives an unconditional estimator of the state vector. However, it is still an optimal estimator in the sense that it minimizes the covariance matrix of the estimation error associated with it.

There are basically three types of smoothing algorithms in a linear model. These are fixed-point smoothing, fixed-lag smoothing and fixed-interval smoothing. Fixed point smoothing involves computing smoothed estimates of the state vector at some fixed point in time. Thus, it gives $\mathbf{a}_{f|t}$ for particular values of f at all time periods $f < t$. Fixed-lag smoothing computes smoothed estimates for a fixed delay, that is $\mathbf{a}_{t-j|t}$, $t-j < t$ for $j = 1, \ldots, M$, where M is some maximum lag. Fixed-interval smoothing is concerned with the computation of the full set of smoothed estimates for a fixed span of data. It is a technique which yields $\mathbf{a}_{t|T}, t = 1, \ldots, T$ and therefore tends to be most widely used for economic and social reasons.

Lastly, any smoothed estimator is based on at least as much information as the corresponding filtered estimator. This means there is sufficient information to make an accurate judgement, which leads to a reduced estimation error. It follows that the covariance matrix of the estimation error of the smoothed estimator is at most that of the corresponding filtered estimator, that is

$$\mathbf{P}_{t|T} \leq \mathbf{P}_t, \quad t = 1, \ldots, T.$$

The smoothed estimator exists if its elements can be estimated with finite covariance of estimation error, that is if $\mathbf{P}_{t|T}$ is bounded. It should be noted that $\mathbf{P}_{t|T}$ is bounded if $\mathbf{P}_t$ is bounded. Thus, $\mathbf{a}_{t|T}$ exists if $\mathbf{a}_t$ exists. The existence of a smoothed estimator is therefore dependent on the existence of a filtered estimator. The converse of this statement is not true. See [Har89] for full details of the mathematical theory of smoothing.

2.5 Example of the Application of Kalman Filter Equations

Here, we illustrate the working principle of the Kalman filter with a numerical example by performing simulations. Suppose the price of a debt instrument is a random constant and we want to estimate it. Let us assume that we can take various observations of the constant, but the observations are corrupted by a 0.1 measurement noise as a result of the fact that our source of observation is not very accurate. In this example, the state space model is governed by the equations [WB06]

$$\alpha_t = \alpha_{t-1} + \eta_t$$

and

$$y_t = \alpha_t + \epsilon_t.$$

The filter equations and parameters corresponding to the above state space model are given by [WB06]

$$\begin{aligned}
\mathbf{a}_{t|t-1} &= \mathbf{a}_{t-1} \\
\mathbf{P}_{t|t-1} &= \mathbf{P}_{t-1} + Q \\
\mathbf{K}_t &= \mathbf{P}_{t|t-1}(\mathbf{P}_{t|t-1} + H_t)^{-1} \\
\mathbf{a}_t &= \mathbf{a}_{t|t-1} + \mathbf{K}_t(y_t - \mathbf{a}_{t|t-1}) \\
\mathbf{P}_t &= (1 - \mathbf{K}_t)\mathbf{P}_{t|t-1}
\end{aligned}$$

Since a small but non-zero value of the variance of the state disturbance gives more flexibility in using the filter, we choose $Q = 0.00001$. Assume that the true value of the random constant has a standard normal distribution with mean 0 and variance 1, then the initial guess for the random constant is 0. Therefore, we set $\mathbf{a}_0 = 0$ before starting the recursion.

Also, let $\mathbf{P}_0$ be the initial value for $\mathbf{P}_{k-1}$. Since we are not absolutely sure that our initial state estimate $\mathbf{a}_0 = 0$, we cannot have $\mathbf{P}_0 = 0$ because if we do, it will cause the filter to initially and always believe that $\mathbf{a}_t = 0$ for every t. As a result, we choose any $\mathbf{P}_0 \neq 0$ and the filter would eventually converge. In particular, let us start with an initial choice $\mathbf{P}_0 = 1$.

2.5.1 Simulations

To begin, we randomly choose a scalar constant 0.37727 as our observation of the price of the debt instrument. We simulate 50 distinct measurements that have errors normally distributed with mean zero and variance 0.01 (recall that observations are corrupted by a 0.1 measurement noise). These measurements allow us to run several simulations with the same measurement noise so that we can make meaningful comparisons between simulations with different parameters.

In the first simulation, we put the measurement error variance at $H = 0.01$ and since this is its true value, we would expect it to perform well in terms of balancing responsiveness and estimate variance. This fact will become more evident in the subsequent simulations. Figure 2.1 shows the results of the first simulation. Clearly, as depicted in the Figure 2.1, the estimate is very close to the true value so it is an optimal estimator.

Recall that when we considered the choice for $\mathbf{P}_0$, we mentioned that the filter would eventually converge so long as $\mathbf{P}_0 \neq 0$. Figures 2.2 and 2.3 show how the value of $\mathbf{P}_t$ changes with time. It settles from

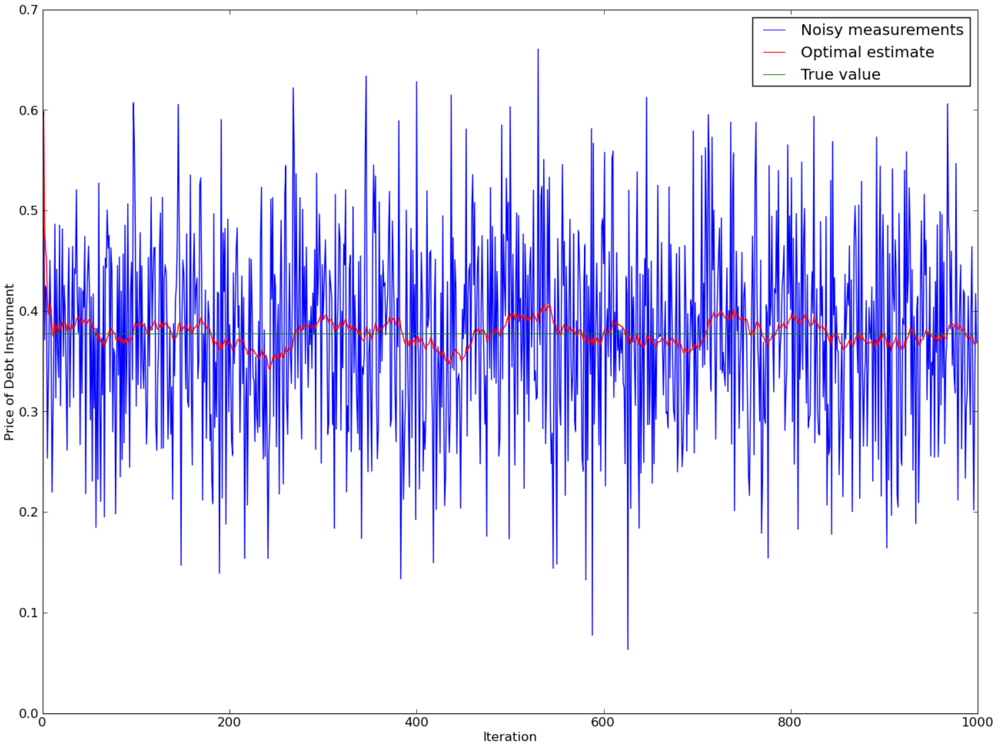

Figure 2.1: Kalman filter simulation with $H = 0.01$.

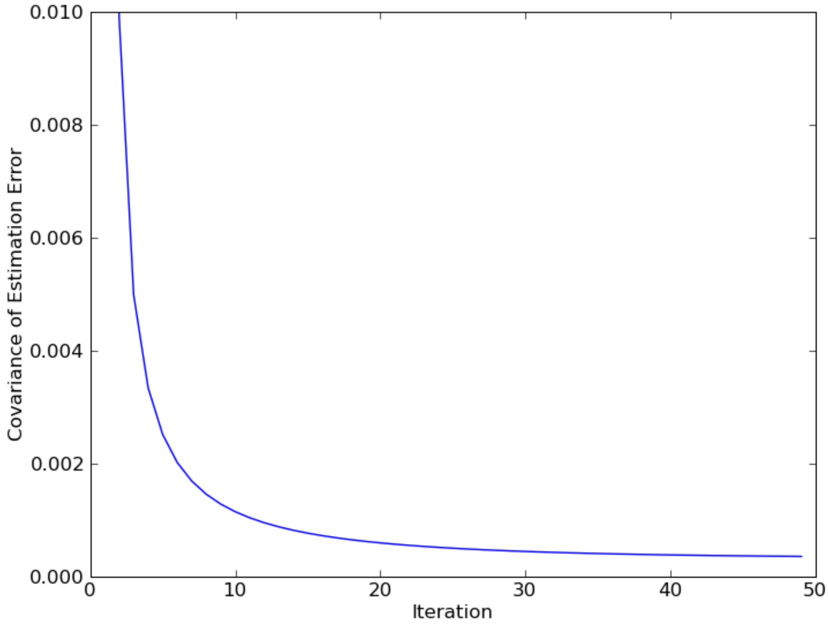

Figure 2.2: The error covariance evolution during the simulation.

our initial and rough choice of 1 to approximately 0.0003411 by the $50th$ iteration and 0.0003113 by the $1000th$ iteration. No matter how much we increase the number of iterations, the value of $\mathbf{P}_t$ can never be zero, but it will attain its smallest possible value. This confirms that the Kalman filter gives optimal estimates that minimize the covariance of the estimation error.

Let us investigate what happens when we change the values of H and Q to obtain different filter performance. Figure 2.4 and Figure 2.5 show the effect when H is increased or decreased by a factor of 100 respectively. In Figure 2.4 the filter was told that measurement noise variance was 100 times greater, with $H = 1$, so it was slower to believe the measurements. This is because the measurements are considered as having a lot of noise and the filter believes more in measurements with small noise, i.e. it believes more in the previous estimation that has small noise.

Figure 2.5 shows that the filter was told that the measurement variance was 100 times smaller, with $H = 0.0001$, so it was very quick to believe the noisy measurements. This is because the measurements are considered as having less noise and the filter believes more in measurements with small noise. As the filter is told that the measurements are very trustful, the Kalman gain is bigger and the estimation follows the path of the measurements. This example clearly demonstrates how the Kalman filter works.

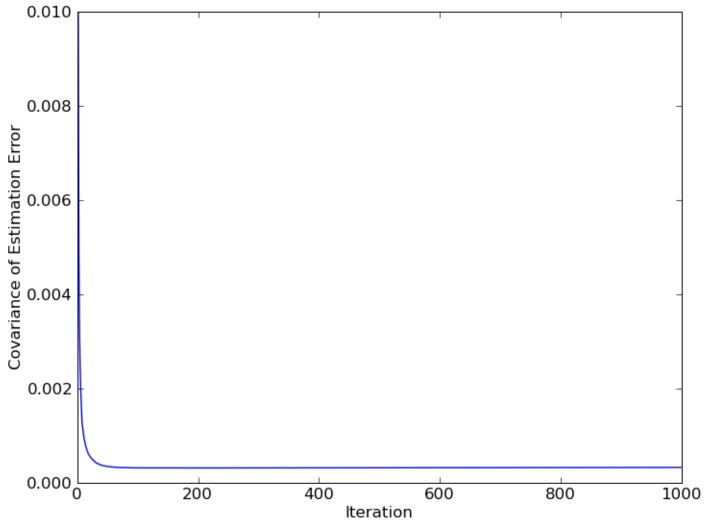

Figure 2.3: There is no notable change in the error covariance for higher number of iterations

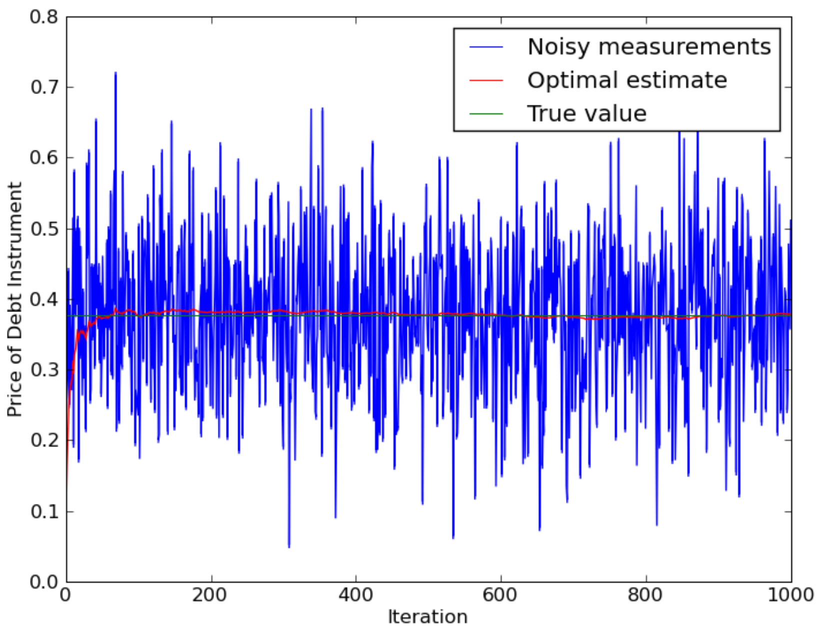

Figure 2.4: Kalman filter simulation with $H = 1$

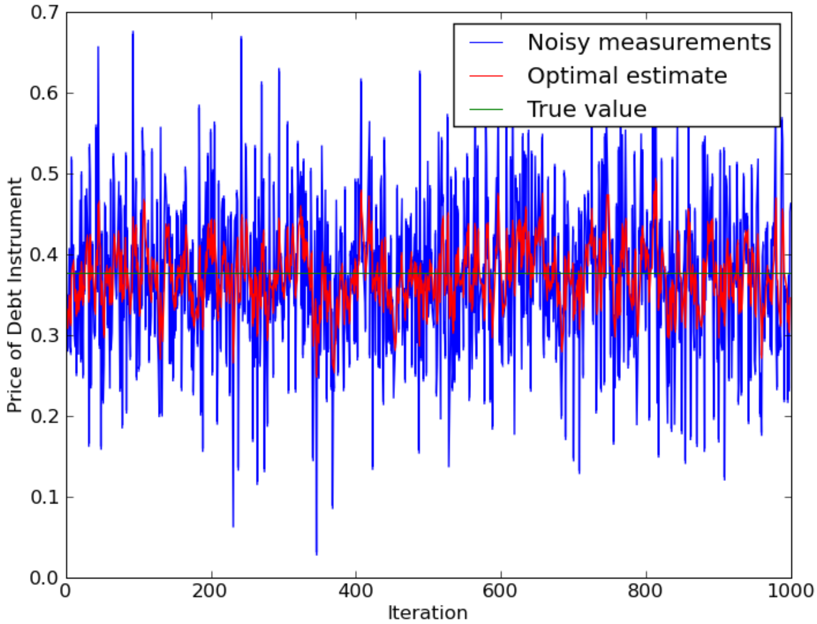

Figure 2.5: Kalman filter simulation with $H = 0.0001$

3. The Generalized Vasicek Term Structure Models

Term structure models are financial models that give the relationship between yields, maturities and prices of bonds. The bonds of interest are the zero coupon bonds, which are otherwise known as pure discount bonds. They are bonds in which no payments of any kind are made on them until maturity. As the alternative name suggests, they are issued below the par value. The par value is the amount that the issuer of the bond is willing to pay the holder at maturity.

There are a number of very popular term structure models in finance. Some of them are

1. Hull-White (extended Vasicek), 1990

$$dr = (\Theta(t) - a(t)r)dt + \sigma(t)dW, \quad a > 0$$

2. Hull-White (extended CIR), 1990

$$dr = (\Theta(t) - a(t)r)dt + \sigma(t)\sqrt{r}dW, \quad a > 0$$

3. Ho-Lee, 1986

$$dr = \Theta(t)dt + \sigma dW,$$

4. Cox-Ingersoll-Ross (CIR), 1985

$$dr = a(b - r)dt + \sigma\sqrt{r}dW, \quad a > 0$$

5. Dothan, 1978

$$dr = ardt + \sigma dW, \quad a > 0$$

6. Vasicek, 1977

$$dr = a(b - r)dt + \sigma dW, \quad a > 0,$$

7. Merton, 1973

$$dr = bdt + \sigma dW.$$

See [Bjo98] for a detailed analysis of these models.

In this chapter, we will study the generalized Vasicek term structure models of interest rates with special focus on both one-and two-factor models.

3.1 The Generalized Vasicek Term Structure Model

One of the earliest stochastic models of the term structure was developed by Vasicek. His model is based on the evolution of an unspecified short-term interest rate.

The generalized Vasicek term structure models are multi-factor mean-reverting Gaussian models of the instantaneous spot interest rate. A possible specification of the instantaneous spot interest rate due to Babbs and Nowman [BN98] and [BN99] is given by

$$r(t) = w_0 + \sum_{j=1}^{J} w_j X_j(t), \tag{3.1}$$

with $w_j = -1$ for $j = 1, \dots J$ and $w_0 = \mu$. So, we have

$$r(t) = \mu - \sum_{j=1}^{J} X_j(t), \tag{3.2}$$

where μ is the long-run mean rate, and $X_1(t), X_2(t), \dots, X_J(t)$ represent the current effects of J streams of economic news whose impact dies away exponentially with time according to the equation

$$dX_j = -\xi_j X_j dt + c_j dW_j, \quad j = 1, 2, \dots, J, \tag{3.3}$$

where $\xi_1, \xi_2, \dots, \xi_J$ are the mean reversion coefficients, $c_1, c_2, \dots, c_J$ are the diffusion coefficients, and $W_1, W_2, \dots, W_J$ are standard Brownian motions with correlation coefficients ρ_{jk}, j, k=$1, \dots, J$. Therefore, (3.3) can be written equivalently as

$$dX_j = -\xi_j X_j dt + \sum_{q=1}^{Q} \kappa_{jq} dZ_q, \quad (Q \leq J), \tag{3.4}$$

where $Z_1, \dots, Z_Q$ are independent standard Brownian motions and κ_{jq} are the associated diffusion coefficients. The parameters κ_{jq}, κ_{kq}, ρ_{jk}, c_j and c_k are related by

$$\sum_{q=1}^{Q} \kappa_{jq} \kappa_{kq} = \rho_{jk} c_j c_k. \tag{3.5}$$

3.1.1 Zero Coupon Bond Price

Bond prices are not in general unique. With different models, we have different bond prices. The price of a zero coupon bond in the generalized Vasicek term structure models is discussed below. It is derived in terms of a finite set of state variables with correlated innovations. The parameters of the models may or may not be constant. However, when they are constant, they provide a very easy way of deriving some very useful results that aid the complete analysis of zero coupon bonds.

Theorem 3.1.1. *Let θ_q be the deterministic market price of risk attaching to each Z_q.[BN99] In the generalized Vasicek term structure model, the time t price of a zero coupon bond with maturity M is*

$$B(M,t) = \exp\left\{ -\int_t^M \mu(u) du - \sum_{q=1}^{Q} \int_t^M \theta_q \sigma_q(M,u) - \frac{1}{2}\sigma_q^2(M,u) du + \sum_{j=1}^{J} \frac{G_j(M) - G_j(t)}{G_j'(t)} X_j(t) \right\} \tag{3.6}$$

where $0 \leq t \leq M$,

$$G_j(t) = \int_0^t \exp\left\{-\int_0^u \xi_j(s)ds\right\} du, \tag{3.7}$$

and

$$\sigma_q(M,t) = \sum_{j=1}^{J} \frac{G_j(M) - G_j(t)}{G_j'(t)} \kappa_{jq}(t), \tag{3.8}$$

μ is the long-run mean rate, θ_q is the market price of risk, $\sigma_q(M,t)$ is the component of the volatility of $B(M,t)$ attributable to Z_q and ξ_j is the mean reversion coefficient.

Proof. The result can be derived by making assumptions on technology and preferences, and restricting information to that generated by the state variables. See [BN99] for a detailed proof. □

3.1.2 Zero Coupon Bond Yield

Consider a zero coupon bond with price $B(M,t)$ as given above. The continuously compounded zero coupon yield $R(M,t)$ is given by

$$R(M,t) = -\frac{\log B(M,t)}{\tau}. \tag{3.9}$$

Hence from (3.6), we see that the zero coupon bond yield in the generalized Vasicek models of the term structure of interest rates is given by

$$R(M,t) = \frac{1}{\tau}\left(\int_t^M \mu(u)du + \sum_{q=1}^{Q}\int_t^M \theta_q\sigma_q(M,u) - \frac{1}{2}\sigma_q^2(M,u)du - \sum_{j=1}^{J}\frac{G_j(M)-G_j(t)}{G_j'(t)}X_j(t)\right). \tag{3.10}$$

3.1.3 Yield on a Zero Coupon Bond with Infinite Maturity

Suppose the maturity of a zero coupon bond is infinite and the parameters remain constant, we want to derive an expression for its yield when the maturity is at infinity. The assumption of constant parameters helps simplify the model and makes it possible to obtain useful results for analysis.

Let $R(\infty)$ represent the yield on a zero coupon bond with infinite maturity. Since $\tau = M - t$ for $t < M$, then $M \to \infty$ implies $\tau \to \infty$. Consequently,

$$R(\infty) = \lim_{M\to\infty} R(M,t) = \lim_{\tau\to\infty} R(\tau + t, t), \tag{3.11}$$

Based on the assumption of constant parameters, (3.10) becomes

$$R(M,t) = \frac{1}{\tau}\left(\int_t^M \mu du + \sum_{q=1}^{Q}\int_t^M \theta_q\sigma_q(M,u) - \frac{1}{2}\sigma_q^2(M,u)du - \sum_{j=1}^{J}\frac{G_j(M)-G_j(t)}{G_j'(t)}X_j(t)\right), \tag{3.12}$$

where

$$\int_t^M \mu du = \mu\tau, \quad \tau = M - t.$$

From (3.7) and (3.8), we have

$$\begin{aligned}\sum_{j=1}^{J}\frac{G_j(M)-G_j(t)}{G_j'(t)}X_j(t) &= \tau\sum_{j=1}^{J}\frac{\left(1-e^{-\xi_j\tau}\right)}{\xi_j\tau}X_j(t)\\ &= \tau\sum_{j=1}^{J}H(\xi_j\tau)X_j(t), \qquad (3.13)\end{aligned}$$

where $H(\xi_j\tau)=\frac{\left(1-e^{-\xi_j\tau}\right)}{\xi_j\tau}$. Furthermore,

$$\begin{aligned}\sigma_q(M,u) &= \sum_{j=1}^{J}\frac{G_j(M)-G_j(u)}{G_j'(u)}\kappa_{jq}\\ &= \sum_{j=1}^{J}\frac{1}{\xi_j}\left(1-e^{-\xi_j(M-u)}\right)\kappa_{jq}\end{aligned}$$

and

$$\sigma_q^2(M,u)=\left(\sum_{j=1}^{J}\frac{1}{\xi_j}\left(1-e^{-\xi_j(M-u)}\right)\kappa_{jq}\right)^2.$$

Integrating $\sigma_q(M,u)$ and $\sigma_q^2(M,u)$ with respect to u from t to M gives

$$\begin{aligned}\int_t^M\theta_q\sigma_q(M,u)du &= \int_t^M\theta_q\sum_{j=1}^{J}\frac{(1-e^{-\xi_j(M-u)})}{\xi_j}k_{jq}du\\ &= \tau\sum_{j=1}^{J}\frac{\theta_q\kappa_{jq}}{\xi_j}\left(1-H(\xi_j\tau)\right)\end{aligned}$$

and

$$\begin{aligned}\int_t^M\sigma_q^2(M,u)du = &\int_t^M\left(\sum_{j=1}^{J}\frac{(1-e^{-\xi_j(M-u)})}{\xi_j}k_{jq}\right)^2du\\ &= \int_t^M\left(\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j}\right)^2du+\int_t^M\left(\sum_{j=1}\frac{\kappa_{jq}e^{-\xi_j(M-u)}}{\xi_j}\right)^2du\\ &-2\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j}\sum_{j=1}^{J}\kappa_{jq}\int_t^M\frac{e^{-\xi_j(M-u)}}{\xi_j}du.\end{aligned}$$

The last expression easily simplifies to

$$\int_t^M\sigma_q^2(M,u)du=\tau\left(\sum_{j=1}^{J}\frac{k_{jq}}{\xi_j}\right)^2+\tau\sum_{i=1}^{J}\sum_{j=1}^{J}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}H\left((\xi_i+\xi_j)\tau\right)-2\tau\sum_{i=1}^{J}\sum_{j=1}^{J}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}H(\xi_j\tau).$$

Consequently,

$$\int_t^M \theta_q\sigma_q(M,u) - \frac{1}{2}\sigma_q^2(M,u) \quad du = \tau\left(\sum_{j=1}^{J}\frac{\theta_q\kappa_{jq}}{\xi_j}\left(1 - H(\xi_j\tau)\right) - \frac{1}{2}\left(\sum_{j=1}^{J}\frac{k_{jq}}{\xi_j}\right)^2\right)$$
$$+\tau\left(-\frac{1}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}H\left((\xi_i+\xi_j)\tau\right) + \sum_{i=1}^{J}\sum_{j=1}^{J}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}H(\xi_j\tau)\right). \tag{3.14}$$

Substituting the above expressions into (3.12) and taking limit as $\tau \to \infty$, we get

$$\lim_{\tau\to\infty} H(\xi_j\tau) = 0$$
$$\lim_{\tau\to\infty} H\left((\xi_i+\xi_j)\tau\right) = 0,$$

and so

$$R(\infty) = \mu + \sum_{q=1}^{Q}\theta_q\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j} - \frac{1}{2}\sum_{q=1}^{Q}\left(\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j}\right)^2. \tag{3.15}$$

This is the expression for the yield on a zero coupon bond with infinite maturity as stated without proof in [BN98].

3.1.4 Constant Parameters Zero Coupon Bond Price

With constant parameters, the zero coupon bond price in (3.6) reduces to the form given in the theorem below.

Theorem 3.1.2. *In the case where the long run average rate μ, the market price of risk processes θ_q, the mean reversion ξ_j and diffusion κ_{jq} coefficients are all constant, the general pricing formula for a pure discount bond evaluates to*

$$B(M,t) = \exp\left\{-\tau\left[R(\infty) - w(\tau) - \sum_{j=1}^{J}H(\xi_j\tau)X_j(t)\right]\right\}, \tag{3.16}$$

with

$$R(\infty) = \mu + \sum_{q=1}^{Q}\theta_q\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j} - \frac{1}{2}\sum_{q=1}^{Q}\left(\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j}\right)^2, \tag{3.17}$$

$$w(\tau) = \sum_{j=1}^{J}H(\xi_j\tau)\left[\sum_{q=1}^{Q}\theta_q\frac{\kappa_{jq}}{\xi_j} - \sum_{q=1}^{Q}\sum_{i=1}^{J}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}\right] + \frac{1}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}H((\xi_i+\xi_j)\tau)\sum_{q=1}^{Q}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}, \tag{3.18}$$

where

$$\tau \equiv M - t$$

and

$$H(\gamma) = \frac{1 - e^{-\gamma}}{\gamma}.$$

Here τ is the term to maturity of the zero coupon bond and H is a function of ξ_j and τ; γ is a dummy variable.

Proof. When the parameters of the model are constant, the general zero coupon bond price reduces to

$$B(M,t) = \exp\left\{ -\int_t^M \mu du - \sum_{q=1}^{Q} \int_t^M \theta_q \sigma_q(M,u) - \frac{1}{2}\sigma_q^2(M,u)du + \sum_{j=1}^{J} \frac{G_j(M) - G_j(t)}{G_j'(t)} X_j(t) \right\}, \tag{3.19}$$

where

$$\begin{aligned} G_j(t) &= \int_0^t \exp\left\{ -\int_0^u \xi_j ds \right\} du \\ &= \frac{1}{\xi_j}(1 - e^{-\xi_j t}), \end{aligned}$$

and

$$\begin{aligned} \sigma_q(M,u) &= \sum_{j=1}^{J} \frac{G_j(M) - G_j(u)}{G_j'(u)} \kappa_{jq} \\ &= \sum_{j=1}^{J} \frac{1}{\xi_j}\left(1 - e^{-\xi_j(M-u)}\right). \end{aligned}$$

Substituting (3.13), (3.14) and the above expressions into (3.19) and simplifying, we obtain

$$B(M,t) = \exp\left\{ -\tau \left[R(\infty) - w(\tau) - \sum_{j=1}^{J} H(\xi_j \tau) X_j(t) \right] \right\} \quad t\epsilon[0,M],$$

which is the required time t price of a pure discount bond with maturity at M, where $R(\infty)$ and $w(\tau)$ are given by (3.17) and (3.18) respectively. $R(\infty)$ represents the yield on a zero coupon bond with infinite maturity as derived in the previous subsection.

In summary, the proof is obtained by assuming the parameters of the model are constant and substituting (3.14) into (3.6). □

3.2 One-Factor Model

When only one factor has an effect on the term structure of interest rates in a particular market, we can use a one-factor model to study many interesting things about the market. A one-factor model assumes that changes in the term structure of interest rates are brought about by one factor. In other words, variations in the interest rates are caused by a single factor. This factor can make interest rates to rise or fall and it is responsible for the changes in the term structure. For a one-factor model, the multi-factor models studied in the previous section assume a special case known as the Vasicek model. In this special case, $J = 1$ and the parameters of the model are constant. Thus, in the special case of a one-factor model, the generalized Vasicek term structure models with constant parameters reduce to the well-known model of Vasicek. The Vasicek model can be thought of as a one-factor model with constant parameters. It is one of the earliest no-arbitrage interest rate models. It gives an explicit equation for the zero coupon yield curve even when the innovations in the state variables are correlated. It can also be used to create interest rate trees. It is based on the idea of mean reverting interest rates where mean reversion speed plays a very important role. Mean reversion is the tendency of a process to revert to its long run mean. If a process is mean reverting, it tends to revert to a constant long run mean. The speed of mean reversion measures the average time it takes for a process to revert to its long run mean. Mean reversion is reasonable for interest rates because it is economically unreasonable to think that interest rates can wander and go to infinity or become arbitrarily huge.

Many term structure models, including the Vasicek model, assume that the fundamental source of uncertainty in the economy is the short rate. The short rate is the annualized rate of return on a very short term investment. All rates ultimately depend on the short term interest rate, which we call the instantaneous spot interest rate and denote by r. The Vasicek model describes the evolution of interest rates and explains the movements of interest rates when they are only driven by market risk. In the Vasicek model, the spot rate defines the term structure and only one factor has an effect on the term structure. The model is also applied to interest rates derivative valuation. The Vasicek model was the first economic model to capture the value of mean reversion. Mean reversion is an important characteristic of interest rates. It is what sets interest rates apart from other financial prices. Hence, interest rates cannot rise indefinitely because such a rise may affect economic activity adversely causing a decrease in interest rates. Also, interest rates cannot decrease indefinitely. No matter how long, they move within a limited range and tend to revert to a long run value, the equilibrium value. In the language of mathematics, we say interest rates are bounded and converge to a real number.

The major disadvantage of the Vasicek model is that interest rates can become negative. The Cox-Ingersoll-Ross model addresses this short coming of the Vasicek model. The Hull-White model further extends the Vasicek model.

3.2.1 Term Structure of a One-Factor Model

As stated earlier, a one-factor model is a special case of the generalized Vasicek models of the term structure of interest rates in which $J = 1$ and the parameters of the model are constant. We shall obtain expressions for the term structure of a one-factor model using results obtained in the previous sections.

The instantaneous spot interest rate is given by

$$r(t) = \mu - X_1(t),$$

where $X_1(t)$ represents the current effect of one factor on the term structure. This factor is an unobservable state variable that can only be estimated. Its impact dies away exponentially according to the stochastic differential equations

$$\begin{aligned} dX_1 &= -\xi_1 X_1 + c_1 dW_1 \\ dX_1 &= -\xi_1 X_1 + \kappa_{11} dZ_1, \quad \kappa_{11} = c_1. \end{aligned}$$

Eliminating X_1 between $r(t)$ and dX_1 gives

$$dr = \xi_1(\mu - r)dt - c_1 dW_1, \tag{3.20}$$

which is a case of the Vasicek model. The negative sign stems from one of the restrictions stated in [BN98].

The time t price of a zero coupon bond with maturity date M in a one-factor model is given by

$$B_1(M,t) = \exp\left\{-\tau\left[R(\infty) - w(\tau) - H(\xi_1\tau)X_1(t)\right]\right\}, \tag{3.21}$$

and the yield is given by

$$R_1(M,t) = \mu + \frac{\theta_1\kappa_{11}}{\xi_1} - \frac{1}{2}\left(\frac{\kappa_{11}}{\xi_1}\right)^2 - w_1(\tau) - H(\xi_1\tau)X_1(t), \tag{3.22}$$

where

$$\begin{aligned} w_1(\tau) &= H(\xi_1\tau)\left[\theta_1\frac{\kappa_{11}}{\xi_1} - \left(\frac{\kappa_{11}}{\xi_1}\right)^2\right] + \frac{1}{2}H(2\xi_1\tau)\left(\frac{\kappa_{11}}{\xi_1}\right)^2, \\ H(\xi_1\tau) &= \frac{\left(1 - e^{-\xi_1\tau}\right)}{\xi_1\tau}, \end{aligned}$$

and the yield with infinite maturity is

$$R_1(\infty) = \mu + \frac{\theta_1\kappa_{11}}{\xi_1} - \frac{1}{2}\left(\frac{\kappa_{11}}{\xi_1}\right)^2.$$

These constitute the term structure of interest rates in the one-factor model. The bond price given by (3.21) together with (3.20) is the well known model of Vasicek.

The full parameter set of the one-factor model is $\{\ \mu,\ X_1(0),\ \theta_1,\ \xi_1,\ \kappa_{11}\ \}$, where $\kappa_{11} = c_1$. $X_1(0)$ represents the initial unobservable state variable.

Simulating the Vasicek Model

All rates in the Vasicek model ultimately depend on the spot interest rate. To simulate this rate, we discretize (3.20). Discretization concerns the process of transferring continuous models and equations into discrete counterparts.

We carry out the discretization by considering changes in the interest rate over a short period Δt. This gives

$$\Delta r(t) = \xi_1\left(\mu - r(t)\right)\Delta t - c_1 Z\sqrt{\Delta t}, \quad \Delta t = t_{i+1} - t_i, \quad 0 = t_0 < t_1 < \ldots < t_m, \tag{3.23}$$

with $Z_1, Z_2, \ldots Z_m$ being independent and identically distributed standard normal random variables [BW98].

Figure 3.1 shows the behaviour of the Vasicek model for $\xi_1 = 5\%$, $\mu = 7\%$, $c_1 = 0.11$ and $\Delta t = 0.0001$.

Running this simulation for 1000 time steps, we observe one of the major weaknesses of the Vasicek model: the interest rates can become negative. This is shown in Figure 3.2.

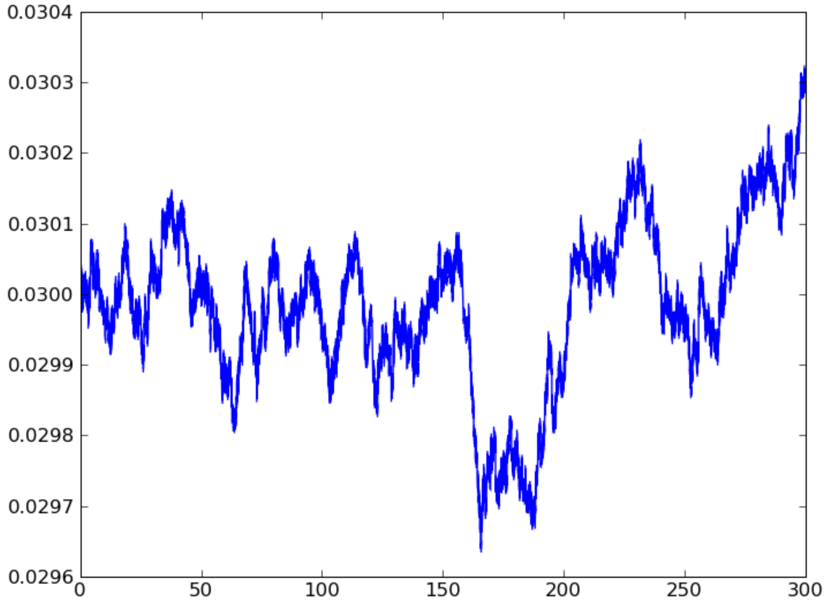

Figure 3.1: Vasicek Simulation for 300 Time Steps

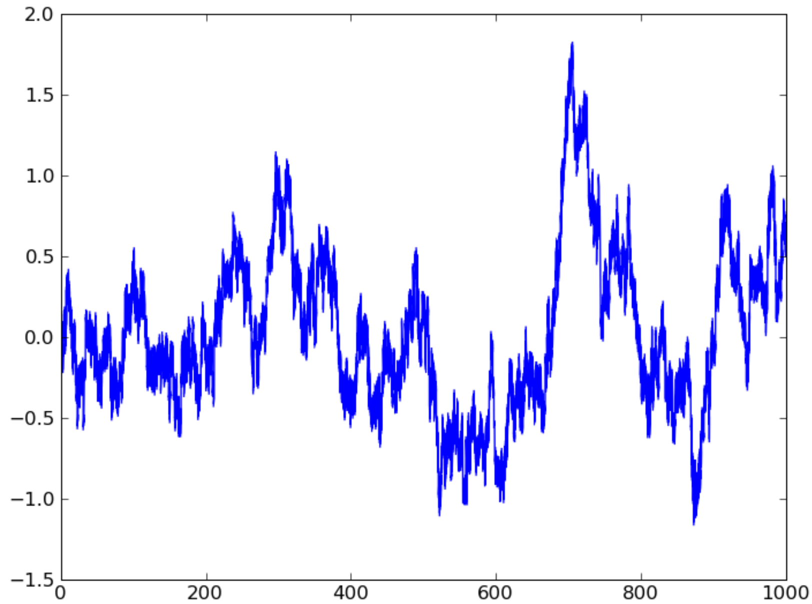

Figure 3.2: Vasicek Simulation for 1000 Time Steps

3.3 Two-Factor Model

An economic model is said to be a two-factor model if it discusses two factors as predominate or exclusive causes of some events. When two unobservable factors cause changes in the term structure of interest rates, we have a two-factor model. The two factors may be correlated and alter the yield curve. If the correlation coefficient between the two factors is not zero, then we say the factors have a gradual impact on the term structure. If the correlation coefficient between the two factors is zero, then we say the two factors have a gradually decaying impact on the term structure. With regard to the generalized Vasicek term structure models being considered in this chapter, we have a two-factor model when $J = 2$. These two factors may be the current effect of two streams of economic news whose impact dies away exponentially. The two factors may be identified with short and long term economic news streams which may be correlated and effect the yield curve. Examples of short term economic news may include rumours of interest rates decision from the Central Bank of a country. Examples of long term economic news may include monthly and quarterly economic statistics [BN98]. These news have a way of effecting the yield curve. They are unobservable and cannot be measured. We can only estimate their impact on the system.

3.3.1 Term Structure of a Two-Factor Model

The term structure of a two-factor model is obtained in the same manner as that of a one-factor model. Everything we want to know about a two-factor model is obtained by setting $J = 2$ in the generalized Vasicek term structure model. With this information, we now proceed to finding the term structure of a two-factor model.

The instantaneous spot interest rate is given by

$$r(t) = \mu - (X_1(t) + X_2(t)), \tag{3.24}$$

where μ is the long run average rate in the two-factor model; $X_1(t)$ and $X_2(t)$ are the two unobservable factors, which cannot be measured and which alter the term structure. As stated earlier, their impact on the term structure can only be estimated. They represent the current effect of two factors, for examples two streams of economic news.

Each news can be modelled as soon as it arrives. The arrival of each type of news is modelled by the innovations of Brownian motions, which may be correlated, while the impact of a piece of news on the term structure wanes away exponentially with time. Consequently,

$$dX_1 = -\xi_1 X_1 dt + c_1 dW_1 \tag{3.25}$$

$$dX_2 = -\xi_2 X_2 dt + c_2 dW_2 \tag{3.26}$$

These equations show how the impact of the short-and long-term economic news streams on the term structure of interest rate dies away exponentially with time. Using matrix notation, we have

$$d\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} -\xi_1 & 0 \\ 0 & -\xi_2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} dt + \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix} d\begin{pmatrix} W_1 \\ W_2 \end{pmatrix}',$$

where the superscript denotes transpose.

Since the news may be correlated, we let $\rho_{12} = \rho_{21} = \rho$ represent the coefficient of correlation, where $\rho \epsilon [-1, 1]$. Therefore, from (3.4), we note that (3.25) and (3.26) can be expressed equivalently as

$$\begin{aligned} dX_1 &= -\xi_1 X_1 dt + \kappa_{11} dZ_1 \\ dX_1 &= -\xi_1 X_1 dt + \kappa_{12} dZ_2 \end{aligned}$$

and

$$\begin{aligned} dX_2 &= -\xi_2 X_2 dt + \kappa_{21} dZ_1 \\ dX_2 &= -\xi_2 X_2 dt + \kappa_{22} dZ_2, \end{aligned}$$

where Z_1, Z_2 are independent standard Brownian motions. From (3.5), we can deduce that κ_{11}, κ_{12}, κ_{21}, κ_{22} are related to c_1, c_2 and the correlation coefficient ρ_{jk} by

$$\kappa_{j1}\kappa_{k1} + \kappa_{j2}\kappa_{k2} = \rho_{jk} c_j c_k, \quad j = 1, 2 \quad k = 1, 2, \tag{3.27}$$

where

$$\rho_{jk} = \begin{cases} \rho & : j \neq k \\ 1 & : j = k. \end{cases}$$

So, each piece of news corresponds to two equations.

The parameters of a two-factor model are all constant. In the light of this, the time t price of a zero coupon bond with maturity at M is

$$B_2(M, t) = \exp\left\{-\tau\left[R(\infty) - w(\tau) - H(\xi_1\tau)X_1(t) - H(\xi_2\tau)X_2(t)\right]\right\},$$

and the yield is given by

$$R_2(M, t) = R_2(\infty) - w_2(\tau) - H(\xi_1\tau)X_1(t) - H(\xi_2\tau)X_2(t)$$

with

$$\begin{aligned} R_2(\infty) &= \mu + \sum_{q=1}^{2} \theta_q \left(\frac{\kappa_{1q}}{\xi_1} + \frac{\kappa_{2q}}{\xi_2}\right) - \frac{1}{2}\sum_{q=1}^{2}\left(\frac{\kappa_{1q}}{\xi_1} + \frac{\kappa_{2q}}{\xi_2}\right)^2 \\ &= \mu + \left(\frac{\kappa_{11}}{\xi_1} + \frac{\kappa_{21}}{\xi_2}\right)\left[\theta_1 - \frac{1}{2}\left(\frac{\kappa_{11}}{\xi_1} + \frac{\kappa_{21}}{\xi_2}\right)\right] + \left(\frac{\kappa_{12}}{\xi_1} + \frac{\kappa_{22}}{\xi_2}\right)\left[\theta_2 - \frac{1}{2}\left(\frac{\kappa_{12}}{\xi_1} + \frac{\kappa_{22}}{\xi_2}\right)\right] \\ &= \mu + \varrho\left[\theta_1 - \frac{1}{2}\varrho\right] + \varsigma\left[\theta_2 - \frac{1}{2}\varsigma\right] \end{aligned}$$

and

$$w_2(\tau) = \sum_{j=1}^{2}\left[H(\xi_j\tau)\left(\theta_q \frac{\kappa_{jq}}{\xi_j} - \zeta - \Lambda\right) + \frac{1}{2}[\zeta H((\xi_1 + \xi_j)\tau) + \Lambda H((\xi_2 + \xi_j)\tau)]\right],$$

where $R_2(\infty)$ is the yield on a zero coupon bond with infinite maturity under a two-factor model and

$$\begin{aligned}
\varrho &= \left(\frac{\kappa_{11}}{\xi_1} + \frac{\kappa_{21}}{\xi_2}\right) \\
\varsigma &= \left(\frac{\kappa_{12}}{\xi_1} + \frac{\kappa_{22}}{\xi_2}\right) \\
\zeta &= \sum_{q=1}^{2} \frac{\kappa_{1q}\kappa_{jq}}{\xi_1 \xi_j} \\
\Lambda &= \sum_{q=1}^{2} \frac{\kappa_{2q}\kappa_{jq}}{\xi_2 \xi_j} \\
H(\gamma) &= \frac{(1 - e^{-\gamma})}{\gamma}.
\end{aligned}$$

We next find explicit expressions for κ_{11}, κ_{12}, κ_{21} and κ_{22} in terms of c_1, c_2 and ρ. Now, from (3.27), the following equations emerge

$$\begin{aligned}
\kappa_{11}^2 + \kappa_{12}^2 &= c_1^2 \\
\kappa_{11}\kappa_{21} + \kappa_{12}\kappa_{22} &= \rho c_1 c_2 \\
\kappa_{21}\kappa_{11} + \kappa_{22}\kappa_{12} &= \rho c_2 c_1 \\
\kappa_{21}^2 + \kappa_{22}^2 &= c_2^2.
\end{aligned}$$

Solving these equations, we get

$$\begin{aligned}
\kappa_{11} &= c_1 \\
\kappa_{12} &= 0 \\
\kappa_{21} &= c_2 \rho \\
\kappa_{22} &= c_2 \sqrt{1 - \rho^2},
\end{aligned}$$

which represent the requisite explicit relationship.

The full parameter set of the two-factor model is

$$\{\mu, X_1(0), X_2(0), \xi_1, \xi_2, \kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22}, \theta_1, \theta_2\},$$

where θ_1 and θ_2 are the market prices of risk; $X_1(0)$ and $X_2(0)$ represent the initial effect of the two factors. They are the two unobservable factors that effect the yield curve.

4. Application of the Kalman Filter Technique to Generalized Vasicek Term Structure Models

We consider the application of the Kalman filter to the generalized Vasicek term structure models. In particular, we consider the application to one- and -two factor interest rate models of the generalized models presented in the previous chapter. We now see how the Kalman filter equations can be used to estimate the unobservable variables and parameters of the generalized models. First, the term structure models are written in a state space form which allows for measurement errors and the Kalman filter equations are then obtained for the term structure models using results in Chapter 2.

4.1 State Space Formulation

This section presents the derivation of the state space formulation of the generalized Vasicek term structure models with special attention to one-and two-factor models. The state space formulation of the interest rate models consists of the measurement and transition equations.

4.1.1 Measurement Equation

The measurement equation contains the observed yields as measured with errors. The measurement errors are additive and normally distributed. We now derive useful expressions that will aid us in properly writing down the measurement equation.

As earlier seen, the theoretical yield on a zero coupon bond in a one-factor model is

$$R_1(t+\tau,t) = R_1(\infty) - w_1(\tau) - H(\xi_1\tau)X_1(t).$$

Similarly, in a two-factor model, it is

$$\begin{aligned} R_2(t+\tau,t) &= R_2(\infty) - w_2(\tau) - [H(\xi_1\tau)X_1(t) + H(\xi_2\tau)X_2(t)] \\ &= R_2(\infty) - w_2(\tau) - \begin{pmatrix} H(\xi_1\tau) & H(\xi_2\tau) \end{pmatrix} \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} \\ &= R_2(\infty) - w_2(\tau) - \begin{pmatrix} H(\xi_1\tau) \\ H(\xi_2\tau) \end{pmatrix}' \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}. \end{aligned}$$

Continuing in this manner, we see that for any $J > 2$, we have

$$\begin{aligned} R(t+\tau,t) &= R(\infty) - w(\tau) - \begin{pmatrix} H(\xi_1\tau) \\ H(\xi_2\tau) \\ \vdots \\ \vdots \\ \vdots \\ H(\xi_J\tau) \end{pmatrix}' \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ \vdots \\ \vdots \\ X_J(t) \end{pmatrix} \\ &= A(\tau) - L(\tau)'X(t), \end{aligned}$$

where the superscript denotes transpose,

and

$$A(\tau) = R(\infty) - w(\tau),\ L(\tau) = \begin{pmatrix} H(\xi_1\tau) \\ H(\xi_2\tau) \\ \vdots \\ \vdots \\ \vdots \\ H(\xi_J\tau) \end{pmatrix} \text{ and } X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ \vdots \\ \vdots \\ X_J(t) \end{pmatrix}.$$

$A(\tau)$ is a scalar and $L(\tau)$ is a $J \times 1$ vector. The scalar $A(\tau)$ and the vector $L(\tau)$ are functions of the term to maturity τ and the parameters of the model. To see this, we recall the expressions of $R(\infty)$, $w(\tau)$ and $H(\xi_j\tau)$ from the previous chapter. Using these expressions, we see that

$$A(\tau) = \mu + \sum_{q=1}^{Q}\theta_q\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j} - \frac{1}{2}\sum_{q=1}^{Q}\left(\sum_{j=1}^{J}\frac{\kappa_{jq}}{\xi_j}\right)^2 -$$
$$\sum_{j=1}^{J}H(\xi_j\tau)\left[\sum_{q=1}^{Q}\theta_q\frac{\kappa_{jq}}{\xi_j} - \sum_{q=1}^{Q}\sum_{i=1}^{J}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}\right] + \frac{1}{2}\sum_{i=1}^{J}\sum_{j=1}^{J}H((\xi_i+\xi_j)\tau)\sum_{q=1}^{Q}\frac{\kappa_{iq}\kappa_{jq}}{\xi_i\xi_j}$$

and

$$L(\tau) = \begin{pmatrix} \frac{1-e^{-\xi_1\tau}}{\xi_1\tau} \\ \frac{1-e^{-\xi_2\tau}}{\xi_2\tau} \\ \vdots \\ \vdots \\ \vdots \\ \frac{1-e^{-\xi_J\tau}}{\xi_J\tau} \end{pmatrix},$$

where μ, θ_q, ξ_j, κ_{jq} are the parameters of the model and τ is the term to maturity. This clearly shows that $A(\tau)$ and $L(\tau)$ are indeed functions of the term to maturity and the parameters of the model.

Suppose there are N observed interest rates $R_1, R_2, \ldots, R_N$ with corresponding terms to maturity $\tau_1, \tau_2, \ldots, \tau_N$ at time t_k, for $k = 1, 2, \ldots, n$. Let R_k denote the set of N interest rates at time t_k corresponding to N terms to maturity, then $R_k = \{R_{1k}, R_{2k}, \ldots, R_{Nk}\} = \{R_{ik} : i = 1, 2, \ldots, N\}$, where

$$\begin{aligned} R_{ik} &= -\frac{\log B(t_k+\tau_i, t_k)}{\tau_i} \\ &= R_{ik}(\infty) - w(\tau_i) - \sum_{j=1}^{J}H(\xi_j\tau_i)X_j(t_k) \end{aligned}$$

A good interpretation of the above notation is to say that at time t_1, which is the first observation date and could be the first month, week or year, there are N observed yields $R_1 = \{R_{11}, R_{21}, \ldots, R_{N1}\}$ with corresponding terms to maturity τ_1, τ_2, ..., τ_N.

With the above expressions clearly spelt out, we proceed to writing down the measurement equation. The measurement equation takes into account the fact that the interest rates are observed or measured with errors. The measurement errors in the interest rates are assumed to be additive and normally distributed. The measurement equation is then given by

$$R_k = Z(\psi)X(t_k) + d(\psi) + \epsilon_k, \quad \epsilon_k \sim N(0, H(\psi)), \tag{4.1}$$

where ψ contains the unknown parameters of the model including the parameters from the distribution of the measurement error. The matrix $d(\psi)$ is an $N \times 1$ matrix whose ith row is given by $A(\tau_i;\psi)$, while $Z(\psi)$ is an $N \times J$ matrix whose ith row is given by $-L(\tau_i;\psi)'$, where $i = 1, 2, \ldots, N$. The notation $A(\tau_i;\psi)$ and $-L(\tau_i,\psi)$ are written to buttress the fact that they are functions of the terms to maturity and the unknown parameters of the model.

The elements of $d(\psi)$ and $Z(\psi)$ are therefore given by

$$d(\psi) = \begin{pmatrix} A(\tau_1;\psi) \\ A(\tau_2;\psi) \\ \vdots \\ \vdots \\ \vdots \\ A(\tau_N;\psi) \end{pmatrix}$$

and

$$Z(\psi) = \begin{pmatrix} -L(\tau_1;\psi)' \\ -L(\tau_2;\psi)' \\ \vdots \\ \vdots \\ \vdots \\ -L(\tau_N;\psi)' \end{pmatrix},$$

where $L(\tau_i;\psi) = \begin{pmatrix} H(\xi_1\tau_i) \\ H(\xi_2\tau_i) \\ \vdots \\ \vdots \\ \vdots \\ H(\xi_J\tau_i) \end{pmatrix}$ and $A(\tau_i;\psi) = R(\infty) - w(\tau_i)$, for $i = 1, \ldots, N$.

Thus

$$Z(\psi) = \begin{pmatrix} -H(\xi_1\tau_1) & -H(\xi_2\tau_1) & \ldots & -H(\xi_J\tau_1) \\ -H(\xi_1\tau_2) & -H(\xi_2\tau_2) & \ldots & -H(\xi_J\tau_2) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ -H(\xi_1\tau_N) & -H(\xi_2\tau_N) & \ldots & -H(\xi_J\tau_N) \end{pmatrix}$$

and

$$d(\psi) = \begin{pmatrix} R(\infty) - w(\tau_1) \\ R(\infty) - w(\tau_2) \\ \vdots \\ \vdots \\ \vdots \\ R(\infty) - w(\tau_N) \end{pmatrix}.$$

This confirms that $Z(\psi)$ is an $N \times J$ matrix and $d(\psi)$ is an $N \times 1$ matrix.

For a one-and a two-factor model, $J = 1$ and $J = 2$. So, $Z(\psi)$ reduces to

$$Z_1(\psi) = \begin{pmatrix} -H(\xi_1\tau_1) \\ -H(\xi_1\tau_2) \\ \vdots \\ \vdots \\ \vdots \\ -H(\xi_1\tau_N) \end{pmatrix} \tag{4.2}$$

and

$$Z_2(\psi) = \begin{pmatrix} -H(\xi_1\tau_1) & -H(\xi_2\tau_1) \\ -H(\xi_1\tau_2) & -H(\xi_2\tau_2) \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ -H(\xi_1\tau_N) & -H(\xi_2\tau_N) \end{pmatrix} \tag{4.3}$$

respectively. Clearly, $d(\psi)$ is not factor dependent and so its order remains the same no matter how many factors we are considering.

The error terms ϵ_k are measurement errors which allow for noise in the sampling process of the data. The variance-covariance matrix of the measurement errors of yields is assumed to have the form $H = h_1, \ldots, h_N$ along the diagonal, where h_1 is the variance of the measurement errors of yields R_{1k} with maturity τ_1 and h_i is the variance of the measurement errors of yields R_{ik} with maturity τ_i, for $i = 1, \ldots, N$ and $k = 1, \ldots, n$.

4.1.2 Transition Equation

The transition equation is represented by the exact discrete-time distribution of the state variables obtained from the solution of (3.3). The exact discrete-time distribution of the state variables is obtained by solving the linear stochastic differential equation

$$dX_j = -\xi_j X_j dt + c_j dW_j, \quad j = 1, 2, \ldots, J.$$

To that end, we let $Y = e^{\xi_j t} X_j$; then

$$\begin{aligned} dY &= \xi_j e^{\xi_j t} X_j dt + e^{\xi_j t} dX_j \\ &= \xi_j e^{\xi_j t} X_j dt + e^{\xi_j t}(-\xi_j X_j dt + c_j dW_j) \\ &= e^{\xi_j t} c_j dW_j, \end{aligned} \tag{4.4}$$

which, upon integrating both sides from t_{k-1} to t_k, gives

$$\begin{aligned} \int_{t_{k-1}}^{t_k} dY &= \int_{t_{k-1}}^{t_k} e^{\xi_j t} c_j dW_j \\ e^{\xi_j t_k} X_j(t_k) - e^{\xi_j t_{k-1}} X_j(t_{k-1}) &= \int_{t_{k-1}}^{t_k} e^{\xi_j t} c_j dW_j \\ X_j(t_k) &= e^{-\xi_j (t_k - t_{k-1})} X_j(t_{k-1}) + c_j e^{-\xi_j t_k} \int_{t_{k-1}}^{t_k} e^{\xi_j t} dW_j \\ &= e^{-\xi_j (t_k - t_{k-1})} X_j(t_{k-1}) + \eta_j(t_k), \end{aligned} \tag{4.5}$$

where

$$\eta_j(t_k) = c_j e^{-\xi_j t_k} \int_{t_{k-1}}^{t_k} e^{\xi_j t} dW_j.$$

For a one-factor model, $j = 1$ and so we have

$$X_1(t_k) = e^{-\xi_1 (t_k - t_{k-1})} X_1(t_{k-1}) + \eta_1(t_k).$$

For a two-factor model, $j = 1, 2$. Thus

$$\begin{aligned} X_1(t_k) = e^{-\xi_1 (t_k - t_{k-1})} X_1(t_{k-1}) + \eta_1(t_k) \\ X_2(t_k) = e^{-\xi_2 (t_k - t_{k-1})} X_2(t_{k-1}) + \eta_2(t_k), \end{aligned}$$

which in matrix form gives

$$\begin{pmatrix} X_1(t_k) \\ X_2(t_k) \end{pmatrix} = \begin{pmatrix} e^{-\xi_1 (t_k - t_{k-1})} & 0 \\ 0 & e^{-\xi_1 (t_k - t_{k-1})} \end{pmatrix} \begin{pmatrix} X_1(t_{k-1}) \\ X_2(t_{k-1}) \end{pmatrix} + \begin{pmatrix} \eta_1(t_k) \\ \eta_2(t_k) \end{pmatrix}.$$

For any $J > 2$, we have

$$\begin{pmatrix} X_1(t_k) \\ X_2(t_k) \\ \vdots \\ \vdots \\ \vdots \\ X_J(t_k) \end{pmatrix} = \begin{pmatrix} e^{-\xi_1 (t_k - t_{k-1})} & 0 & \dots & \dots & 0 \\ 0 & e^{-\xi_2 (t_k - t_{k-1})} & 0 & \dots & 0 \\ \vdots & 0 & e^{-\xi_3 (t_k - t_{k-1})} & 0 & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & 0 & \\ 0 & 0 & 0 & \dots & e^{-\xi_J (t_k - t_{k-1})} \end{pmatrix} \begin{pmatrix} X_1(t_{k-1}) \\ X_2(t_{k-1}) \\ \vdots \\ \vdots \\ \vdots \\ X_J(t_{k-1}) \end{pmatrix} + \begin{pmatrix} \eta_1(t_k) \\ \eta_2(t_k) \\ \vdots \\ \vdots \\ \vdots \\ \eta_J(t_k) \end{pmatrix},$$

thus,

$$X(t_k) = \Phi(\psi)X(t_{k-1}) + \eta(t_k).$$

By change of notation, we have

$$X_k = \Phi(\psi)X_{k-1} + \eta_k, \tag{4.6}$$

which is the required transition equation and is a VAR(1) model. As seen above, $\Phi(\psi) = e^{-\xi_j(t_k - t_{k-1})}$ for $j = 1, 2, \ldots, J$ generates the above matrix and η_k is normally distributed with $E(\eta_k) = 0$ and $Cov(\eta_k) = V(\psi)$.

The measurement and transition equations so obtained represent the state space formulation of our model and open the way for us to present the set of Kalman filter equations associated with the generalized Vasicek term structure model.

4.2 Kalman Filtering of Generalized Vasicek Term Structure Models

Now that the models have been successfully put in a state space form, the Kalman filter recursive equations can be employed to estimate the unobserved state variables and the parameters of the models. In this section, we will apply the Kalman filter equations discussed in the previous chapter to the generalized Vasicek term structure models. The main aim of the Kalman filter equations is to obtain information about X_k from the observed interest rates in the measurement equation. The Kalman filter equations also help estimate the unknown parameters of the models. The Kalman filter gives an optimal estimator of X_k based on the available information from the observed yields. The optimal estimator of X_k based on the available information up to a certain time is the conditional expectation of X_k given all the available information up to that time. The Kalman filter recursive equations can be broken down into prediction and update step.

4.2.1 Prediction Equations

Prediction involves estimating the unobserved state variables at a particular time based on the available information up to the time before that particular time. This information comes from the observed variables. In this case, the observed variables are the observed interest rates. So, the information comes from the observed interest rates.

Let $E_{k-1}(X_k)$ denote the conditional expectation of the unknown state vector X_k given the available information up to time t_{k-1} and let $\hat{X}_{k/k-1}$ denote the optimal estimator of X_k based on the observed interest rates up to time t_{k-1}. The optimal estimator of X_k based on the observed interest rates up to time t_{k-1} is the conditional expectation of X_k given the available information up to time t_{k-1}. As a result, we have

$$\begin{aligned} \hat{X}_{k|k-1} &= E_{k-1}(X_k) \\ &= E_{k-1}(\Phi(\psi)X_{k-1} + \eta_k) \\ &= \Phi(\psi)E_{k-1}(X_{k-1}) + E_{k-1}(\eta_k) \\ &= \Phi(\psi)\hat{X}_{k-1}, \end{aligned} \tag{4.7}$$

where $E_{k-1}(X_{k-1}) = \hat{X}_{k-1}$ and $E_{k-1}(\eta_k) = 0$. We next consider the error involved in estimating X_k during the prediction step.

The prediction estimation error is the error connected with estimating X_k based on the observed interest rates up to time t_{k-1}. We are interested in the covariance matrix of the estimation error. Let $\mathbf{P}_{k|k-1}$ represent the covariance matrix of the estimation error. Using the result developed in (2.5), we have that

$$\begin{aligned}\mathbf{P}_{k|k-1} &= E_{k-1}\left[(X_k - \hat{X}_{k|k-1})(X_k - \hat{X}_{k|k-1})'\right] \\ &= \Phi(\psi)E_{k-1}\left[(X_k - \hat{X}_{k-1})(X_k - \hat{X}_{k-1})'\right]\Phi'(\psi) + E_{k-1}\left[\eta_k\eta_k'\right] \\ &= \Phi(\psi)\mathbf{P}_{k-1}\Phi'(\psi) + V, \end{aligned} \tag{4.8}$$

where $E_{k-1}\left[\eta_k\eta_k'\right] = V$ is the so called covariance matrix of η_k.

These two equations are known as the prediction equations. They are used in the prediction step of the Kalman filter algorithm.

4.2.2 Updating Equations

While the prediction step involves obtaining information about X_k from the observed interest rates up to time t_{k-1}, the update step involves obtaining information about X_k based on the observed interest rates up to time t_k. The update step uses additional information on the interest rates R_k at time t_k to obtain a more precise and updated estimator of X_k This new and current estimate of X_k is called the filtered estimate. Once the new observed interest rates become available, $\hat{X}_{k|k-1}$ can be updated to give $\hat{X}_k$. The updating equations also comprise the optimal estimator and the covariance matrix of the estimation error.

Let $\hat{X}_k$ denote the optimal estimator of X_k based on the available information from the observed interest rates up to time t_k and let P_k be the covariance matrix of the estimation error. The updating equations are given by

$$\begin{aligned}\hat{X}_k &= E_k(X_k) \\ &= \hat{X}_{k|k-1} + \mathbf{P}_{k|k-1}Z'F_k^{-1}v_k, \end{aligned} \tag{4.9}$$

and

$$\begin{aligned}\mathbf{P}_k &= E_k[(X_k - \hat{X}_k)(X_k - \hat{X}_k)'] \\ &= \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}Z'F_k^{-1}Z\mathbf{P}_{k|k-1} \\ &= \left(\mathbf{P}_{k|k-1}^{-1} + Z'H^{-1}Z\right)^{-1}, \end{aligned} \tag{4.10}$$

where

$$\begin{aligned} v_k &= R_k - (d + Z\hat{X}_{k|k-1}) \\ F_k &= H + Z\mathbf{P}_{k|k-1}Z'. \end{aligned}$$

We now justify the expression written in (4.10). Eliminating F_k from $\mathbf{P}_k$, we have

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}Z'(H + Z\mathbf{P}_{k|k-1}Z')^{-1}Z\mathbf{P}_{k|k-1}.$$

Using the fact that $HH^{-1} = I$ and $(EG)^{-1} = G^{-1}E^{-1}$, we obtain

$$\begin{aligned}(H + Z\mathbf{P}_{k|k-1}Z')^{-1} &= \left(H(I + H^{-1}Z\mathbf{P}_{k|k-1}Z')\right)^{-1}\\ &= (I + H^{-1}Z\mathbf{P}_{k|k-1}Z')^{-1}H^{-1}.\end{aligned}$$

Hence,

$$\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}Z'(I + H^{-1}Z\mathbf{P}_{k|k-1}Z')^{-1}H^{-1}Z\mathbf{P}_{k|k-1}.$$

The generalized inverse formula for the sum of invertible matrices, [HS81]

$$A - AU(I + BVAU)^{-1}BVA = (A^{-1} + UBV)^{-1}, \tag{4.11}$$

gives

$$\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}Z'(I + H^{-1}Z\mathbf{P}_{k|k-1}Z')^{-1}H^{-1}Z\mathbf{P}_{k|k-1} = (\mathbf{P}^{-1}_{k|k-1} + Z'H^{-1}Z)^{-1},$$

thus,

$$\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}Z'F_k^{-1}Z\mathbf{P}_{k|k-1} = (\mathbf{P}^{-1}_{k|k-1} + Z'H^{-1}Z)^{-1}$$

Consequently, (4.10) follows directly.

We have thus obtained the prediction and updating equations. These equations absolutely specify the Kalman filter. They form the complete Kalman filter technique used for optimally estimating the unobservable state variables and the parameters of the generalized Vasicek models of the term structure of interest rates.

5. Conclusion

We have studied the term structure of interest rate models and have presented the Kalman filter equations. In particular, we have studied the generalized Vasicek term structure models. We also presented the application of the Kalman filter to the generalized Vasicek term structure models of zero coupon bonds in an introductory way. Building of the term structure models of zero coupon bonds is an extremely complex area in financial mathematics and all we have done is to focus on a subclass of models which have been tested. That is why we have considered the generalized Vasicek models. Over time, the generalized Vasicek models have been particularly preferable in interest rate modelling. This is because they are analytically tractable in the sense that the pure discount bond prices can be related to the state variables by a closed-form formula even when the parameters are not constant and innovations in the state variables are not uncorrelated. This contrasts with other models, for example the multi-factor CIR models, where there are no closed-form formulas for the price of a zero coupon bond unless innovations in the state variables are uncorrelated. This confers additional flexibility and superiority on every Gaussian model; generalized Vasicek models are a subclass of the Gaussian models. The major weakness, however, lies in the fact that interest rates can become negative as seen in the simulation performed in the course of this essay. It is for this liability that other models, for example, the CIR models, surpass the generalized Vasicek models and are particularly designed to address this major shortcoming of the generalized Vasicek models. The use of the state space formulation and the application of the Kalman filter to the generalized Vasicek term structure models have the advantage that they allow the underlying state variables to be handled correctly as unobservable state variables. They also present a good way of optimally estimating noisy and unobservable interest rates.

As the literature on the subject of interest rate modelling and Kalman filter is vast and expanding, there is definitely a plethora of things to learn, especially with respect to the Kalman filtering of interest rate models applied to real world markets. Although building of term structure models is undoubtedly mathematically rigorous, my ultimate objective in the near future is to construct term structure models with universal acceptance that will address many of the deficiencies in other known term structure models. As there are very many factors that impact the term structure of interest rates in the real world, another stimulating area of my future research will be to apply the Kalman filter techniques to cases where the generalized Vasicek term structure models assume higher factors, say $J > 2$, and implement the techniques in emerging markets such as the Nigerian and the South African Markets. These are areas of application where the generalized Vasicek models have not yet been fully explored. The aim will then be to see whether the models can represent the term structure in those markets.

Acknowledgements

Father God Osanobua, who has given me the grace to be alive, I give you praise well well O. Na persin wey dey alive na him fit aspire to achieve something, no be so?

I would like to thank my supervisor, Emeritus Professor Ronnie Becker, for proposing this very stimulating topic and for guiding me to its successful completion. It is my dream to continue doing research in this area. Many thanks are due to my big brothers, Eimienwanlan Ibhagui and Oseiwe Ibhagui, my big sisters, Ighogboya Ibhagui, Onuwahagbe Ibhagui, Ester Oladapo and Patricia Ibhagui, and my parents, Mr. Lawrence and Mrs. Regina Ibhagui, for their care, love, support and constant prayers. My close pals, Segun Ojo and Oderinde Temitope Rabiu, have been very helpful to me. The role Segun Ojo played in ensuring my coming to study at AIMS will forever be appreciated. Oderinde Temitope Rabiu has always been there for me both morally and materially. I thank Akintona Samuel Olawale, Joel Ummuna, Ngozi Onyeagbazi and Maitami for their unstinting support during our days in Akoka. I am indebted to one of the finest previous AIMS students, Wole Solana, from whom I indirectly got the information that led to my applying to AIMS. I thankfully recognize the role that Mr. Obinna Nwosu and Dr. Alfred Dele Damilare played in making me follow the path that has led me to where I am today. I am grateful to Dimbinaina Ramarimbahoaka for fixing one of my Python codes and to Andry Rabenantoandro and Andriamihaja Ramanantoanina for working with me in proving some of the results presented in this essay. The extremely witty IT manager, Mr. Jan Groenewald, has made my computing skills to grow in leaps and bounds, and the adroit English language teacher, Frances Aron, has helped polish my ability in the queen's language. Both of you are very much held in the highest possible esteem. I am also grateful to my extremely vivacious tutor, Trust Chibawara, for his constant constructive criticism ($\mathbf{c}^3$) of my essay. This has made the finished work to approximate perfection. I have also benefited considerably from discussions with other tutors and resident researchers at AIMS, and I am particularly grateful for the advice I have received from Dr. Madeleine Cule of the University of Oxford. My conversations with Gaston Kuzamunu Mazandu have been especially valuable in providing me with insight into the various ways my research capacity can be developed and in convincing me that patience in research is indeed the best way to proceed. Outside AIMS, I have gained enormously from contacts with Dr. Niyi Isifiade of the University of Cape Town, Dr. Albert Melareri of the University of Cape Town Teaching Hospital, Dr. Akinyeye of the University of the Western Cape, Brother Paul and all other members of the Redeemed Christian Church of God, Latter House Paris. What is more, I would like to express my deepest gratitude to fellow Nigerian students comprising two kings, Solomon and Olufemi, two queens, Wunmi and Tolu, and one senior queen, Chika, as well as ex-students of AIMS, especially Victoria, Chidinma, Felix and Bode, for their incredibly wondrous enunciations and no strings attached guidance during our wee days in South Africa. There are no better compatriots elsewhere!

All of my colleagues, from other African countries, who have made my ten months at AIMS an exciting and intellectually stimulating experience, I say you are all forever blessed. I will continue to value your inestimable friendships, which even measure theory has failed to accurately fathom. Finally and importantly, I say a very big thank you to one of the best orators that the world has ever known, a leader in no small measure, a senior lion in the den of great lions and an initiator of life changing ideas, I say thank you to Professor Niel Turok for founding AIMS.

References

[Bjo98] Tomas Bjork, *Arbitrage theory in Continuous time*, Oxford University Press, 1998.

[BN98] Simon H. Babbs and K. Ben Nowman, *An application of generalized Vasicek term structure models to the UK gilt-edged market: a Kalman filtering analysis*, Applied Financial Economics **8** (1998), 637–644.

[BN99] Simon H. Babbs and K. Nowman, *Kalman filtering of generalized vasicek term structure models*, Journal of Financial and Quantitative Analysis **34** (1999), 115–130.

[BW98] Simmon Benninga and Zvi K. Wiener, *Term Structure of Interest Rates*, 1998, Available from `http://pluto.mscc.huji.ac.il/mswiener/zvi.html`.

[CZ03] Marek Capinski and Tomasz Zastawniak, *Financial Engineering*, Prentice Hall of India, 2003.

[Gar06] Antonio Herrero Garcia, *Interest Rate Model Calibration Using Kalma Filtering*, Available from www.iit.upcomillas.es/pfc/resumenes/449f89ad85750.pdf, 2006.

[Har89] Andrew C. Harvey, *Forecasting, structural time series models and the Kalman filter*, Cambrigde University Press, 1989.

[HS81] H. V. Henderson and S. R. Searle, *On Deriving the Inverse of a Sum of Matrices*, Society for Industrial and Applied Mathematics **23** (1981), 0036–1445.

[MB03] John F. Marshall and Vipul K. Bansal, *Financial Engineering*, Prentice Hall of India, 2003.

[WB06] Greg Welch and Garry Bishop, *An Introduction to the Kalman Filter*, Available from `http://www. cs.unc.edu/ welch`, 2006.